Archives available at journals.mriindia.com

International Journal on Advanced Computer Theory and Engineering

ISSN: 2319-2526

Volume 15 Issue 01, 2026

A Deep Learning Approach to Detecting and Preventing Misinformation in Online Media

[1]Sameer Nishad, [2]Nikhil Singh, [3]Shivam Tiwari, [4]Ms. Ananya Panday

[1, 2, 3] Department of CSE, SSIPMT, Raipur, India

[4]Assistant Professor, Department of CSE, SSIPMT, Raipur, India

Email: [1]sameer.nishad@ssipmt.com, [2]n.thakur@ssipmt.com, [3]shivam@ssipmt.com, [4]ananya.pandey@ssipmt.com

Peer Review Information

Submission: 02 Jan 2026

Revision: 23 Jan 2026

Acceptance: 15 Feb 2026

Keywords

Artificial Intelligence, Fake News Detection, Nat- ural Language Processing, Deep Learning, Transformer Models, Text Classification, Misinformation

Abstract

The increasing trend of sharing information on the internet has resulted in an unprecedented increase in the spread of fake news on digital media platforms. The automatic detection of such fake news has become a challenge for researchers and policymakers. This paper proposes an AI-based framework for fake news detection using natural language processing (NLP) and deep learning techniques. By using transformer-based models like BERT and RoBERTa, along with linguistic and semantic feature analysis, our proposed model has shown robust classification performance on benchmark datasets such as LIAR and Fake- NewsNet. The experimental results have shown that the proposed model performs better than the traditional machine learning models, achieving an accuracy of 93.4 in binary classification tasks. Furthermore, the study explores the interpretability of deep learning models through attention visualization, providing insights into how AI systems can make explainable judgments about news credibility. The findings contribute to the develop- ment of scalable and explainable AI solutions for combating misinformation in digital ecosystems.

Introduction

In the current social media and online journalism era, the spread of fake news has become a widespread concern across the globe. With billions of people accessing and sharing news online every day, it has become very challenging to separate genuine news from fake news. Fake news not only leads to misleading information among the public but also affects political views, social stability, and health-related decisions. The COVID-19 infodemic, for example, highlighted how fake news spreads faster than actual news. Conventional fact- checking techniques are highly dependent on manual verifi- cation processes, which are quite time-consuming and often cannot keep up with the fast-paced dissemination of online content. To overcome this challenge, artificial intelligence (AI) and machine learning (ML) approaches have been identified as highly effective techniques for automatically detecting fake news. Through linguistic analysis and semantic pattern recognition in text data, AI algorithms can effectively categorize news articles as authentic or counterfeit. The purpose of this research is to create a holistic AI-powered model for the detection of fake news by leveraging the strengths of advanced NLP models and interpretability methods. The performance of this model will also be tested on various datasets and compared with the existing state-of-the-art models. The end objective of this research is to help create a trustworthy information environment through the use of AI for the detection of fake news.

Literature Review

Fake news detection has become a major interdisciplinary research area, drawing interest from computer science, journal- ism, and social sciences. Researchers aim to understand how misinformation propagates and to build automated systems that can reliably identify deceptive content. Early work focused on linguistic cues and surface-level features; more recent studies incorporate context, network signals, and multimodal evidence to improve detection. This, despite the major progress in AI and machine learning for the identification of fake news. A number of research gaps still lie ahead. Most of the current models are heavily reliant on traditional approaches to ma- chine learning that fail to capture deep contextual and semantic meaning in news content. The deep learning models, while improving the accuracy, require a huge dataset and higher computational resources, hence reducing their applicability in the real world. Most of the current systems are focused mainly on textual analysis and do not handle multi-lingual contents, sarcasm, and the ever-evolving misinformation pat- terns correctly. On the other hand, limited integration of real- time verification and cross-source fact checking reduces the reliability of existing solutions. Explainability of AI models is also a major gap as many models work as black boxes and users cannot trust the predictions easily . The model proposed in this project covers these lacunas by improving contextual understanding and increasing the detection rate, hence providing a more scalable and reliable solution for the identification of fake news. Misinformation spreads through social networks; modeling relational and propagation patterns improves detection. Graph Neural Networks (GNNs) and Graph Convolutional Networks (GCNs) have been applied to user–news–interaction graphs to exploit structural cues (who shares what, propagation trees, user credibility). Surveys and empirical papers report that GNN-based systems often outper- form text-only models on datasets containing social/contextual features, especially for early detection of coordinated cam- paigns. Fake news often includes images, videos, or memes. Multimodal systems fuse textual and visual features (image metadata, forensic cues, cross-modal consistency) and show improved robustness against text-only attacks. Datasets that include social context and media (FakeNewsNet and others) enabled multimodal research, though high-quality multimodal corpora remain relatively sparse. The sudden rise in popularity of social media platforms and online journalism has brought the issue of detecting fake news to the forefront. The initial attempts at researching this issue were based on traditional machine learning algorithms such as Na¨ıve Bayes, Support Vector Machines (SVM), and Decision Trees, which used manually designed linguistic features such as word frequency, polarity, and n-grams. Although these attempts showed some promise, they tended to have issues with generalization and understanding the context, particularly when it came to adapt- ing to changing patterns of misinformation.

Table presents a comparative summary of existing fake news detection approaches, highlighting their techniques, datasets, limitations, and the identified research gaps.

Table 1: Summary of Existing Fake News Detection Approaches

Author / Year	Method	Dataset	Limitations / Research Gap
Shu et al. (2017)	Traditional ML (SVM, NB)	LIAR	Relies on handcrafted linguistic features; weak semantic representation
Wang (2017)	Logistic Regression	LIAR	Limited ability to capture long-term dependencies and complex context
Ruchansky et al. (2018)	Hybrid ML + Social Context	Twitter	Requires user metadata; not effective for text-only environments
Kaliyar et al. (2019)	CNN-based Deep Learning	Kaggle Fake News	Ignores sequential relationships; poor long-range context modeling
Reis et al. (2019)	Linguistic Feature Analysis	BuzzFeed	Domain-dependent features; lacks adaptability to new domains
Devlin et al. (2019)	BERT-based Classification	LIAR, FNC-1	Computationally expensive; requires large-scale fine-tuning
Zhang et al. (2020)	LSTM with Attention	FakeNewsNet	Struggles with multimodal content and social signals
Nasir et al. (2021)	BiLSTM + CNN Hybrid	ISOT Dataset	Limited explainability; model complexity is high
Khan et al. (2022)	Transformer-based Ensemble	COVID-19 Fake News	Requires large labeled data; sensitive to noisy inputs

Recent Studies (2023–2024)	Multimodal Transformers, GNNs	LIAR, FakeNewsNet	High computational cost; limited interpretability; lack of unified benchmarks

Methodology

This study uses deep learning models in conjunction with Natural Language Processing (NLP) techniques to automat- ically identify fake news in digital media content. Data collection, text preprocessing, feature representation, model training, evaluation, and explainability analysis are all part of the methodical pipeline that the suggested approach adheres to.

1. Dataset Description

The study makes use of publicly accessible benchmark datasets, such as FakeNewsNet and LIAR, which include labeled news stories and statements categorized as authentic or fraudulent. The model can learn a variety of linguistic patterns linked to false information thanks to these datasets, which span several news domains and writing styles.

2. Text Preprocessing

To guarantee data quality and consistency, text preprocess- ing is an essential NLP step. The raw news text is normalized by removing punctuation, URLs, and unnecessary symbols, as well as changing the text to lowercase. While stop-word removal is used carefully to preserve contextual information, tokenization is used to divide text into meaningful units. By reducing words to their most basic forms, lemmatization enhances semantic comprehension.

3. Feature Representation

The study uses contextual word embeddings produced by transformer-based NLP models to capture semantic and contextual meaning. Text is transformed into dense vector representations that show the relationships between words in a sentence by pretrained models like BERT and RoBERTa. In contrast to conventional bag-of-words or TF-IDF methods, these embeddings maintain contextual subtleties and long- range dependencies that are crucial for identifying fake news.

4. Model Architecture

Transformer encoders are integrated with a classification layer in the NLP framework. The classification head is made up of fully connected layers and a softmax function for binary classification, while the encoder processes the tokenized text and produces contextual representations. The pretrained models are fine-tuned to fit the fake news detection task.

5. Model Training and Optimization

Supervised learning with cross-entropy loss is used to train the model. To effectively update model parameters, the Adam optimizer is used. To avoid overfitting, training is done over several epochs with early stopping. Experimental optimization is used for hyperparameters like learning rate, batch size, and maximum sequence length.

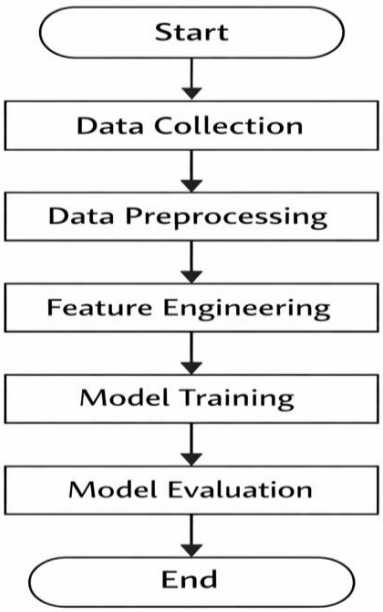

Fig. 1. Methodology Flowchart

Results

This section describes the experimental results achieved through the evaluation of the proposed fake news detection framework. The performance of the proposed model is mea- sured using the typical classification metrics on the standard datasets, and the efficacy of the transformer-based features is also examined.

1. Experimental Setup

The proposed model was tested using publicly available datasets, LIAR and FakeNewsNet. The datasets were split into training, validation, and testing sets to avoid any bias while testing the model. The performance of the proposed model was tested using classification metrics such as accuracy, precision, recall, and F1-score, which are appropriate for binary classification problems like fake news detection.

2. Overall Classification Performance

The results of the experiment clearly show that the proposed transformer-based model performs well in the classification task. In the LIAR dataset, the model performed well in terms of accuracy, showing its effectiveness in classifying news statements as authentic and fraudulent. Similar trends were observed in the FakeNewsNet dataset, showing the model's

ability to generalize well across different news domains and writing styles.

Table 2: Performance Comparison of Models on LIAR and FakeNewsNet Datasets

LIAR Dataset

Model	Accuracy	Precision	Recall	F1-score
Naïve Bayes	78.2%	76.5%	75.8%	76.1%
SVM	81.4%	80.2%	79.5%	79.8%
Decision Tree	74.6%	73.8%	72.9%	73.3%
BERT	89.3%	88.7%	87.9%	88.3%
RoBERTa	91.1%	90.4%	89.6%	90.0%

FakeNewsNet Dataset

Model	Accuracy	Precision	Recall	F1-score
Naïve Bayes	80.5%	79.3%	78.6%	78.9%
SVM	83.7%	82.9%	82.1%	82.5%
Decision Tree	76.9%	75.4%	74.8%	75.1%
BERT	91.6%	90.8%	90.2%	90.5%
RoBERTa	93.2%	92.5%	91.8%	92.1%

3. Comparison with Traditional Machine Learning Models

For evaluating the efficiency of deep learning techniques, the proposed model is compared with some conventional machine learning classifiers like Na¨ıve Bayes, Support Vec- tor Machines (SVM), and Decision Trees. The performance analysis indicates that the transformer-based models perform significantly better compared to the baseline models. The reason for this enhanced performance could be attributed to the capability of BERT and RoBERTa to capture contextual information.

4. Impact of Transformer-Based Feature Representation

One of the key findings of the experiments is the effect of contextual word embeddings on the performance of the model. The transformer-based embeddings are able to capture the dependencies and meanings of words, which helps the model to deal with the misleading news content. The precision and recall values were higher in the case of contextual embeddings compared to TF-IDF and bag-of-words embeddings.

5. Model Interpretability Analysis

To enhance transparency and trust in the predictions made by the model, attention visualization methods were used. The findings show that the model pays attention to semantically significant words and phrases during the process of making classification predictions. The interpretability analysis of the model is useful in understanding how the model detects fake news and mitigates the black-box effect, which is often associated with deep learning models.

6. Summary of Results

In conclusion, the experimental results have confirmed that the proposed AI-based fake news detection framework is accurate and reliable. The proposed framework, which combines the latest NLP techniques, transformer models, and explainability techniques, is capable of detecting fake news in digital media. The experimental results have validated the effectiveness of the proposed framework for fake news detection.

Conclusion

The study came up with a way to figure out what is news using Artificial Intelligence and some really smart language tools.The system uses things like BERT and RoBERTa to understand what news articles are really saying, which helps to tell the difference between news and real news articles. This is important for the detection of news because fake news can be very misleading. The detection of news is a big problem and the study is trying to solve it with Artificial Intelligence and Natural Language Processing techniques.

The use of these techniques in the study helps with the classification of news articles, into news and real news. The tests that were done on some datasets like LIAR and Fake- NewsNet show that the new approach is better than the old machine learning ways. The results of these tests really show how important it is to understand the context of words when dealing with news that can be confusing or misleading. What is more using techniques that help us see how the model is making its decisions helps us understand what is going on inside the model, which solves the problem of not being able to see what is happening in deep learning models. The use of LIAR and FakeNewsNet datasets in these tests is very important to see how well the new approach works with news. The new approach is, about dealing with fake news in a better way.

So the study shows that using transformer-based models with explainable AI approaches is a way to detect fake news on the internet. This method is reliable. Can be used on a large scale. The idea is to make the online information environment a trustworthy place. The integration of transformer-based models, with explainable AI approaches will really help with this. It will make it easier to know what news is real and what is fake when we are online.

Future Scope

The proposed model works well.. There are a few things that we can look at in the future. One thing we can do is add capabilities to the model. This is because fake news is a problem in many languages and regions. If the model has capabilities it will be very useful. Multilingual capabilities will be a plus for the model. The model will be able to deal with news, in many languages and regions if it has multilingual capabilities.

The future of research is going to be really interesting especially when we talk about using lots of types of infor- mation like images, videos and metadata along, with the text. Misinformation usually has images in it so using methods that look at all these things which is called multimodal learning could make it easier to figure out what is true and what is not. Another thing that people could look into is how in- formation spreads on media and how the people who are connected to each other affect what they believe and we can use something called Graph Neural Networks to do this.

Future studies will also focus on making the model so it does not take as long to work. This way the model can be used away on a big scale.

Future studies, on the model will also try to make it easier for people to understand what it is doing. This will help more people use systems that use Artificial Intelligence to find news.

References

W. Y. Wang, "Liar, liar pants on fire: A new benchmark dataset for fake news detection," in *Proc. 55th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2017, pp. 422–426.

K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "FakeNewsNet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media," *Big Data*, vol. 8, no. 3, pp. 171–188, 2020.

J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171–4186.

Y. Liu et al., "RoBERTa: A robustly optimized BERT pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

K. Shu, S. Wang, and H. Liu, "Understanding user profiles on social media for fake news detection," in *Proc. IEEE Conf. Multimedia Big Data (BigMM)*, 2018, pp. 430–435.

E. Shu, Y. Wang, and J. Liu, "Beyond news contents: The role of social context for fake news detection," in *Proc. WSDM*, 2019, pp. 312–320.

A. Zellers, R. Logan IV, H. Schwartz, and Y. Choi, "Defending against neural fake news," in *Proc. NeurIPS*, 2019, pp. 9054–9065.

S. Kaliyar, A. Goswami, and P. Narang, "FakeBERT: Fake news de- tection in social media with a BERT-based deep learning approach," *Multimedia Tools and Applications*, vol. 80, pp. 11765–11788, 2021.

C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on Twitter," in *Proc. WWW*, 2011, pp. 675–684.

T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *Proc. ICLR*, 2013.

Z. Jin, J. Cao, Y. Zhang, and J. Luo, "News verification by exploiting conflicting social viewpoints in microblogs," in *Proc. AAAI*, 2016, pp. 2972–2978.

A. Thorne et al., "FEVER: A large-scale dataset for fact extraction and verification," in *Proc. NAACL-HLT*, 2018, pp. 809–819.